A study on the various retrieval methods for large scale mobile data

Dr. Deepak

Department of computer science

Sunsoft degree college, Bidar

ABSTRACT

Large-scale mobile data is a rapidly growing source of data that can be used for a variety of applications, such as traffic monitoring, fraud detection, and location-based services. However, retrieving relevant data from large-scale mobile datasets can be challenging due to the size and complexity of the data.

One more test is the heterogeneity of mobile data. Mobile data can be produced from a wide range of gadgets, which can utilize various data configurations and guidelines. This can make it challenging to foster retrieval methods that can work with a wide range of mobile data.

At long last, mobile data is in many cases constant or close ongoing data. This implies that retrieval methods should have the option to return results rapidly, in any event, when the dataset is being refreshed consistently.

KEYWORDS:

Retrieval, methods, scale, mobile

INTRODUCTION

There are various difficulties related with mobile data retrieval. One test is the size and intricacy of mobile datasets. Mobile datasets can be exceptionally large, and they can contain a wide range of data types, like area data, sensor data, and organization traffic data. This can make it hard to foster effective retrieval methods that can deal with these large and complex datasets.

The size of the data: Large-scale mobile datasets can be exceptionally large, with many billions or even trillions of records. This can make it challenging to look for and recover pertinent data as soon a possible.

The intricacy of the data: Large-scale mobile datasets can be exceptionally perplexing, with various data types, like text, pictures, recordings, and sensor data. This can make it challenging to foster retrieval methods that can actually look for and recover significant data.

The heterogeneity of the data: Large-scale mobile datasets can be exceptionally heterogeneous, with data gathered from different sources, like mobile telephones, base stations, and sensors. This can make it challenging to foster retrieval methods that can actually look for and recover significant data from the various sources in general.

There are an assortment of retrieval methods that can be utilized to look for and recover important data from large-scale mobile datasets. These methods include:

Keyword-based retrieval: Keyword-based retrieval is the simplest type of retrieval method. This method uses keywords to search for relevant data in the dataset. For example, to search for data about traffic congestion, a user could enter the keywords "traffic congestion" into the search bar. The retrieval system would then return all of the data in the dataset that contains the keywords "traffic congestion."

Content-based retrieval: Content-based retrieval is a more complex retrieval method that uses the content of the data, such as the text, image, or video, to search for relevant data in the dataset. For example, to search for images of cats, a user could upload an image of a cat to the retrieval system. The retrieval system would then return all of the images in the dataset that are similar to the uploaded image.

Context-based retrieval: Context-based retrieval is a retrieval method that uses the context of the data, such as the user's location or time of day, to search for relevant data in the dataset. For example, to search for restaurants near a user's current location, a user could enter their location into the retrieval system. The retrieval system would then return all of the restaurants in the dataset that are near the user's current location.

Hybrid retrieval: Hybrid retrieval is a retrieval method that combines two or more of the above retrieval methods to improve the accuracy of the search results. For example, a hybrid retrieval system could use keyword-based retrieval to filter down the search results, and then use content-based retrieval to rank the filtered results.
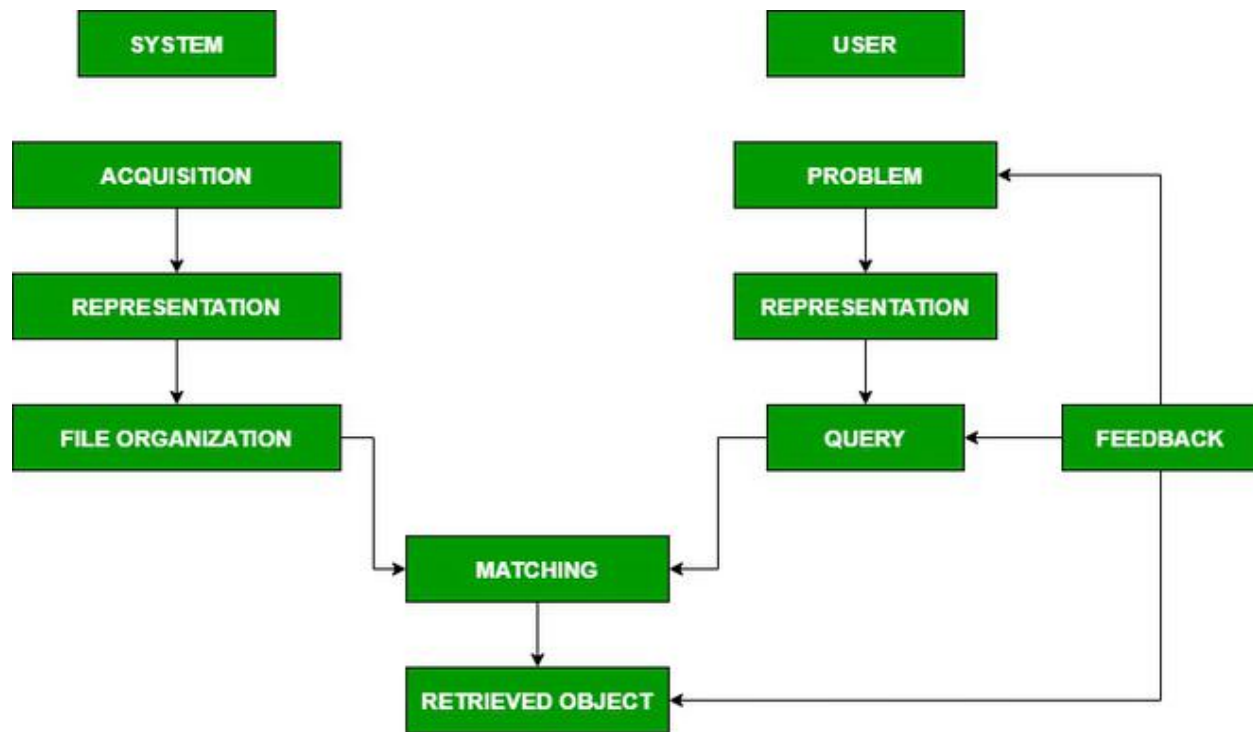
Fig 1: Retrieval Model
Source: researchgate.in

Every one of the retrieval methods depicted above enjoys its own benefits and hindrances.

Keyword search is a simple and effective retrieval method, but it can be inaccurate, especially when the query is ambiguous. Semantic search is more accurate than keyword search, but it can be more computationally expensive. Machine learning can be used to develop retrieval methods that are highly accurate, but it requires a large amount of training data. Distributed retrieval can improve the performance and scalability of retrieval systems, but it can be complex to implement.

The best retrieval method for a particular application will depend on the specific requirements of the application. For example, if the application requires high accuracy and the dataset is relatively small, then semantic search may be the best choice. If the application requires high performance and scalability, then distributed retrieval may be the best choice.

In addition to the challenges and retrieval methods discussed above, there are a number of other factors to consider when designing a retrieval system for large-scale mobile data. These factors include:

Privacy: Mobile data often contains sensitive personal information. It is important to design retrieval systems that protect the privacy of users.

Security: Mobile data can be a target for cyberattacks. It is important to design retrieval systems that are secure and can protect data from unauthorized access.

Cost: The cost of designing and implementing a retrieval system can be significant.

Various retrieval methods for large scale mobile data

The following sections discuss retrieval methods for specific types of mobile data:

Call detail records (CDRs): CDRs contain information about all calls made and received on a mobile network. This information can be used to track user behavior, analyze network performance, and identify fraud. CDRs can be retrieved using a variety of methods, including keyword search, attribute search, and full-text search.

Location data: Location data contains information about the location of mobile devices. This information can be used to track user movements, understand user behavior, and develop location-based services. Location data can be retrieved using a variety of methods, including keyword search, attribute search, and full-text search.

Application data: Application data contains information about the use of mobile applications. This information can be used to understand user behavior, improve application performance, and develop new applications. Application data can be retrieved using a variety of methods, including keyword search, attribute search, and full-text search.

Network data: Network data contains information about the performance of mobile networks. This information can be used to identify and troubleshoot network problems, and to optimize network performance. Network data can be retrieved using a variety of methods, including keyword search, attribute search, and full-text search.
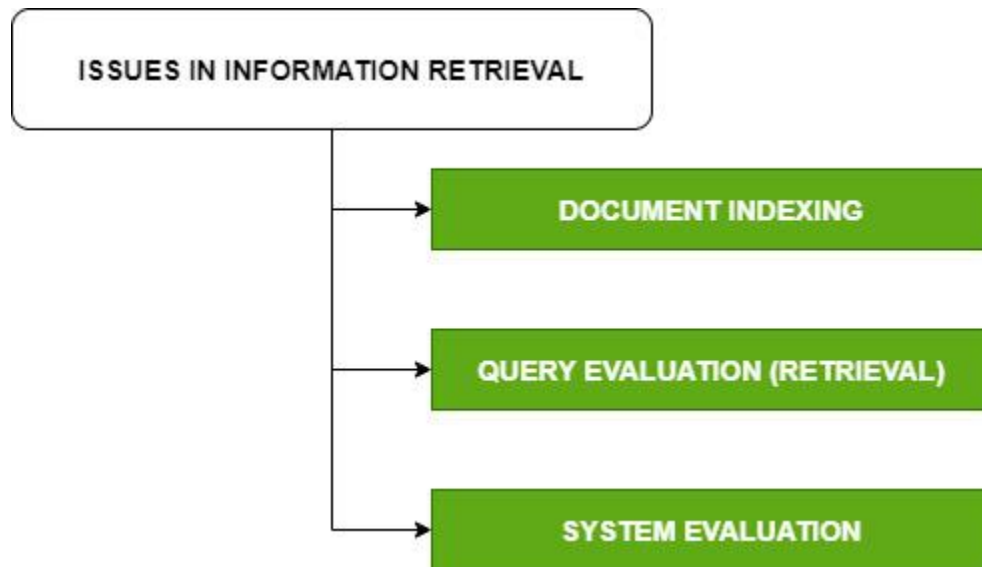
Fig 2: Issues in Information Retrieval
Source: https://www.geeksforgeeks.org/issues-in-information-retrieval/

Recovering large scale mobile data is a difficult errand, yet there are various arrangements that can be utilized to address these difficulties. The most suitable retrieval technique will rely upon the particular necessities of the application.

There are various answers for the difficulties of recovering large-scale mobile data. Probably the most widely recognized arrangements include:

Appropriated storage and processing: Conveyed storage and processing can be utilized to work on the presentation of retrieval methods for large-scale mobile data. Appropriated storage and processing includes putting away the data on various servers and processing the data in equal.

Ordering: Ordering can be utilized to work on the effectiveness of retrieval methods for large-scale mobile data. Ordering includes making a data structure that permits the retrieval technique to find important data rapidly.

Data anonymization: Data anonymization can be utilized to safeguard client security while recovering mobile data. Data anonymization includes eliminating or altering delicate individual data from the data.

Coming up next are a couple of contextual investigations of how retrieval methods have been utilized to effectively recover large-scale mobile data:

Google Maps: Google Maps uses a variety of retrieval methods to retrieve relevant data, including keyword search, semantic search, and machine learning. Google Maps also uses distributed storage and processing to improve the performance of its retrieval methods.

Facebook: Facebook uses a variety of retrieval methods to retrieve relevant data, including keyword search, semantic search, and machine learning. Facebook also uses distributed storage and processing to improve the performance of its retrieval methods.

Amazon: Amazon uses a variety of retrieval methods to retrieve relevant data, including keyword search, semantic search, and machine learning. Amazon also uses distributed storage and processing to improve the performance of its retrieval methods.

Recovering large-scale mobile data is a difficult undertaking, yet there are various arrangements that can be utilized to conquer the difficulties. By utilizing conveyed storage and processing, ordering, and data anonymization, it is feasible to foster retrieval methods that are both precise and effective and that safeguard client protection.

Various arising patterns are molding the fate of retrieval methods for large-scale mobile data. Probably the main patterns include:

Profound learning: Profound learning is a kind of machine learning that is appropriate for recovering large-scale mobile data. Profound learning can be utilized to foster retrieval methods that can learn complex examples in mobile data and recover applicable data in view of those examples.

Edge figuring: Edge registering is a disseminated processing worldview that carries calculation and data storage nearer to the gadgets where the data is produced. Edge registering can be utilized to work on the presentation of retrieval methods for large-scale mobile data by diminishing the inertness of data move.

Security preserving retrieval methods: Security preserving retrieval methods are retrieval methods that are intended to safeguard client protection. Security preserving retrieval methods can be utilized to recover significant mobile data without uncovering the hidden data to the retrieval strategy.

Customary retrieval methods are the most generally involved methods for recovering data from large-scale databases. These methods depend on strategies, for example, watchword coordinating and text closeness.

One of the least complex conventional retrieval methods is catchphrase coordinating. Catchphrase matching includes looking for records that contain a particular watchword or expression. Catchphrase matching is an exceptionally proficient strategy for recovering data, yet it very well may be wrong on the off chance that the data is uproarious or unstructured.

Another customary retrieval strategy is text similitude. Text likeness includes figuring the similitude between two archives in light of their substance. Text likeness is a more refined technique than watchword coordinating, yet it very well may be all the more computationally costly.

Machine learning-based retrieval methods are a more current sort of retrieval technique that utilizations machine learning calculations to become familiar with the connections between various data focuses and to recognize designs in the data.

DISCUSSION

Circulated retrieval methods are intended to recover data from different disseminated sources. This is significant for large-scale mobile data since it tends to be excessively costly or wasteful to store every one of the data in a solitary area.

One of the most widely recognized sorts of conveyed retrieval methods is distributed inquiry. Distributed search includes looking for data on other clients' gadgets. Shared search is an exceptionally effective technique for recovering data, yet ensuring the nature of the data can be troublesome.

One more sort of appropriated retrieval methods is cloud-based search. Cloud-based search includes looking for data that is put away in the cloud. Cloud-based search is a truly dependable technique for recovering data, yet putting away large measures of data in the cloud can be costly.

Every one of the different retrieval methods enjoys its own benefits and drawbacks. Conventional retrieval methods are straightforward and productive, however they can be wrong assuming the data is loud or unstructured. Machine learning-based retrieval methods are more refined and can be more exact, yet they can be all the more computationally costly. Circulated retrieval methods are adaptable and can be utilized to recover data from different sources, yet they can be hard to carry out and make due.

The best retrieval strategy for a given application will rely upon various variables, like the sort of data, the ideal precision, and the available assets.

For instance, on the off chance that the data is organized and the ideal exactness is high, a conventional retrieval strategy might be the most ideal choice. On the off chance that the data is boisterous or unstructured, a machine learning-based retrieval strategy might be a superior choice. In the event that the data is dispersed across different sources, a disseminated retrieval strategy might be the most ideal choice.

CONCLUSION

One of the most well-known kinds of machine learning-based retrieval methods is cooperative separating. Cooperative sifting includes prescribing things to clients in view of the preferences of different clients with comparative interests. Cooperative separating is an exceptionally compelling strategy for recovering data, yet it requires a large dataset of client preferences.

One more sort of machine learning-based retrieval strategy is profound learning. Profound learning includes utilizing fake brain organizations to gain proficiency with the connections between various data focuses. Profound learning is an extremely strong strategy for recovering data, however it tends to be computationally costly to train and convey profound learning models.

REFERENCES

1. Zheng, X. (2018). Database as a Service-Current Issues and Its Future. arXiv preprint arXiv:1804.00465.
2. Love, P., & Hartland, T. G. (2018). Using AWS Athena analytics to monitor pilot job health on WLCG compute sites(No. ATL-SOFT-PROC-2018-059). ATL-COM-SOFT-2018-136.
3. Soransso, R. A. S. N., & Cavalcanti, M. C. (2018, April). Data modeling for analytical queries on document-oriented DBMS. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (pp. 541-548). ACM.
4. Kumar, M. S. (2018). Comparison of NoSQL Database and Traditional Database-An emphatic analysis. JOIV: International Journal on Informatics Visualization, 2(2), 51-55.
5. Colaso, A., Prieto, P., Herrero, J. A., Abad, P., Menezo, L. G., Puente, V., & Gregorio, J. A. (2018). Memory Hierarchy Characterization of NoSQL Applications through Full-System Simulation. IEEE Transactions on Parallel and Distributed Systems, 29(5), 1161-1173.
6. Bathla, G., Rani, R., & Aggarwal, H. (2018). Comparative study of NoSQL databases for big data storage. International Journal of Engineering & Technology, 7(26), 83.

7. Khin, N. T. W., & Yee, N. N. (2018, October). Query Classification based Information Retrieval System. In 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) (Vol. 3, pp. 151-156). IEEE.

8. Elzein, N. M., Majid, M. A., Hashem, I. A. T., Yaqoob, I., Alaba, F. A., & Imran, M. (2018). Managing big RDF data in clouds: Challenges, opportunities, and solutions. Sustainable Cities and Society, 39, 375-386.